Automated Multi-Label Skin Lesion Classification Using Transfer Learning and Advanced Data Augmentation

1st Dr K Venkatesan
Dept. of Computer Science and Engg.
Amrita School of Computing
Amrita Vishwa Vidyapeetham
Chennai,India

2nd Dr K Deepak
Dept. of Computer Science and Engg.
Amrita School of Computing
Amrita Vishwa Vidyapeetham
Chennai,India

3rd A Sohan Avaneesh Guptha
Department of CSE(AI)
Amrita School of Computing
Amrita Vishwa Vidyapeetham
Chennai,India

4th M Hemanth Kumar Reddy
Department of CSE(AI)
Amrita School of Computing
Amrita Vishwa Vidyapeetham
Chennai,India

5th R Nikhileswar Reddy
Department of CSE(AI)
Amrita School of Computing
Amrita Vishwa Vidyapeetham
Chennai,India

6th V V N Uday Aditya
Department of CSE(AI)
Amrita School of Computing
Amrita Vishwa Vidyapeetham
Chennai,India

***Abstract*—Skin cancer is a dangerous disease that can spread to across parts of the skin and often resulting from extended exposure to ultraviolet (UV) rays. The disease, on the other hand, can be cured when detected at an early stage, and now, Computer-aided detection (CAD) and machine learning (ML) have emerged as great resources for dermatologists. The aim of this research was to provide a deep learning solution to the multiclass classification of skin lesions using images from the ISIC2018 dataset, which holds seven lesion types. Considering the class imbalance problem other image augmentation methods were applied as well, including rotation, zooming and shifting to pre-process the images in the training set. The study focused on classifying lesion types using the EfficientNetB3 model, which is favorable for performance while being efficient. The model was trained with a categorical cross-entropy loss function and an Adamax optimizer achieving a high training accuracy of 99.95%, validation accuracy test accuracy of 94.03% 93.46% respectively. Precision, recall and F1-scores were evaluated for all lesion types; the best classification performance was achieved for Basal Cell Carcinoma and Dermatofibroma, while melanoma and nevi had achieved less accuracy. The EfficientNetB3 in combination with data augmentation and optimization strategies performs well on the classification of skin lesions. However, due to the performance of the model on the more severe lesion types such as melanoma, it requires enhancement. These classes will be prioritized in future recovery efforts in order to excellence the responsiveness and recall metrics of the model, so that the model can be more dependable for use in a clinical setting and for early skin cancer detection.**

***Index Terms*—skin cancer, Augmentation, EfficientNetB3, Adamax optimizer, Deep learning**

I. INTRODUCTION

Skin cancer is the prevalent type of cancer in the population and it is constantly detected on every corner of the earth with aggressive lethality. skin lesions are classified into two general categories, primary and secondary. Primary skin lesions are abnormal skin conditions that may be congenital or develop over time. Secondary skin lesions arise from changes or aggravations of primary lesions. In more severe cases,especially with melanoma, cancer cells can break away from the original tumor and spread to other parts of the body like the lymph nodes, lungs, liver or brain, causing life-threatening complications [15]. It has the capacity to increase the risk of developing other type of cancer in future due to weekend immune system. Melanoma is among the worst types of skin cancer with existence chance rate of 5-6% [11]. Melanoma are responsible for around 0.3% of all skin cancer cases in india with over 1,00,000 people 0.16% death. According to world Health Organization (WHO), skin cancer represents one-third of all cancer diagnosed anually, with over 1,30,000 new cases of melanoma and 3 million cases of non-melanoma reported globally each year [12]. There were 1,86,680 new cases of melanoma diagnoised, with 89,070 noninvasive and 97,160 invasive, and 7,990 deaths from the disease. The survival rate of skin cancer patients, especially those diagnosed with melanoma, is said to be higher when the cancer is detected in its early stages [5]. However, lesions such as basal cell carcinoma (BCC) and melanoma are considerably different from one another and have different growth patterns that can affect the treatment and prognosis. However, standard methods of diagnosis such as visual examination or dermoscopy are observer dependent and prone to variability amongst dermatologists [20]. To improve more accuracy in less time and assist less experienced dermatologists, various deep learning based computerized techniques have been developed by computer vision researchers for the detection of skin cancer. Deep learning techniques have made it possible to automate detection processes and in so doing enhance diagnosis by reducing common point of failure that is, human input for a more accurate diagnostic procedure. [17]

The introduction of modeling routines that are powered by a machine learning model and are used in classification of skin lesion has its unique challenges primarily,

Data Imbalance,in a number of datasets that pertain lesional skin, it can be observed that benign lesions like melanocytic nevi, have a higher prevalence than malignant lesions. The model becomes biased towards the majority classes, leading to high overall accuracy but low sensitivity or recall for minority classes like melanoma, which are clinically crucial. The disparity creates bias in the model where it tends to favor common lesions discriminately while neglecting rare but clinically important lesions. Inter-Class and Intra-Class Visual Similarities, with different types of lesion can be visually similar or on the other hand identical types of lesions can have different appearances in terms. For example, melanoma and nevi can have similar morphologies and the degree of heterogeneity that exists within some classes of BCC for instance will make its classification rather difficult. Dataset Variability, differences in image acquisition parameters including lighting, skin color and even the angle of the dermoscopic images also increase the level of challenge in the detection and classification task. Dermoscopic images often contain artifacts such as hair, bubbles, or shadows, which can confuse the model and degrade future extraction [12]. To remove this air bubbles and all we will use data preprocessing. Use pre-trained segmentation models to isolate the lesion region and remove surrounding artifacts. Besides, lesions can also be covered by hair or other elements thus making the classification task even more difficult. Overfitting may occur sometimes by performing well on training dataset but fail to recognize on unseen validation or test data, leading to poor real-world performances [8].

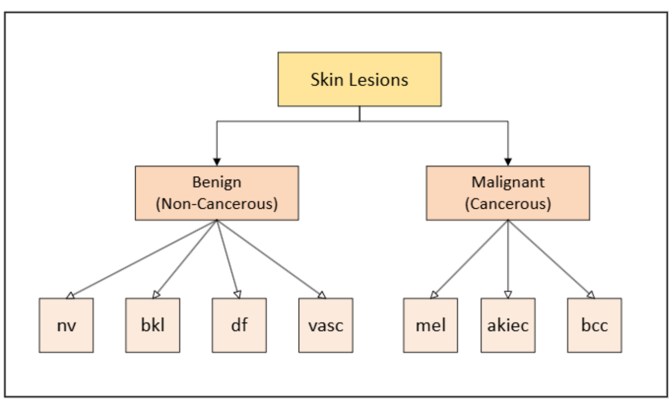

Fig. 1. Distribution of Skin Lesions

The seven categoreis of skin lesions are classified as benign or malignant. Benign lesions include Nevus (nv), which is the generic term for moles-sometimes benign growths within the skin that vary by color, shape, size [13]. Most are harmless. Other benign lesions include dermatofibroma (df)-small, firm nodules that are usually brown or tan and asymptomatic-and benign keratosis (bkl) or non-cancerous growths that might resemble scaly or wart-like patches, such as those called seborrheic keratosis [15]. There are Vascular Lesions (vasc), which are often associated with blood vessels-those being benign tumors as seen in hemangiomas. On the contrary, Malignant lesions include Melanoma (mel), serious and aggressive skin cancer coming from melanocytes. The lesion usually presents itself in form of an irregular mole or new dark spot; Basal Cell Carcinoma (bcc) is the most similar form of skin cancer; usually appearing as a shiny bump or sore that doesn't heal. It grows slowly but does require treatment; and Actinic Keratosis (akiec), which is a precancerous condition characterized by rough, scaly patches from sun damage, which may progress to squamous cell carcinoma if not treated. This classification would help to use it towards estimating the risk whether it is cancer causing or non-cancerous and driving proper clinical management and treatment [3].

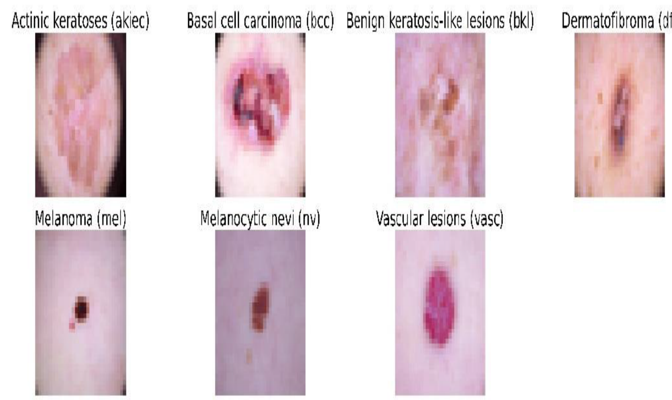

Fig. 2. Distribution of Skin Lesions

EfficientNetB3 is a part of the EfficientNet family, which is known for having a compound scaling method that simultaneously scales the width, depth and resolution of the networks thereby enhancing the accuracy of the model while utilizing fewer computational resources in the process as compared to conventional CNN's [14]. This paper actually uses ISIC2018 dataset which includes images of seven different skin lesion types and it required the robust technique to overcome so we took a pre-trained model of EfficientNetB3 and used it to train for skin lesion classification tasks [12]. This architecture makes it easier to perform calculations on most parts of the medical images that are high in resolution like the dermoscopy images. The following steps were fundamental in ensuring that the model was well optimized: Data preprocessing first and foremost step to get a de-noised image and hair artifacts removal. In this data preprocessing step we will use Blackhat Morphological Filtering and Inpainting is used to remove hair artifacts. other morphological models like erosion or dilation might not specifically target small dark regions. so, we used Blackhat which highlight small, dark regions on a lighter background, which can be useful in skin lesion detection [12]. Traditional models like simple blurring or interpolation can lead to unrealistic or poorly reconstructed areas,which might confuse the model. Inpainting used to reconstruct missing or occluded parts of the image, such as regions covered by hair or other artifacts that can interfere lesion analysis. Transfer Learning, able to leverage general image recognition of our skin lesion dataset thanks to the pre-trained weights on ImageNet dataset. This enhanced the convergence speed and accuracy, even when there is limited labeled data available. Data Augmentation, in order to eliminate the issue of data imbalance and enhance the diversity of training samples, various strategies were applied such as random rotation, flipping, zooming, changing the brightness of the image and shifting its width and height etc. These augmentations enhance the generalization and the robustness of the model as it trains on different image distortions [17]. Some models trained on a specific dataset like ISIC2018 may not generalize well to images from different sources, especially they have different resolution, lightining conditions or noise levels. Poor generalization, can result in decreased accuracy when deploying the model in real-world clinical settings. By using EfficientNetB3 with proper classification techniques we can solve this problems. [19]

EfficientNetB3 has a large number of layers and parameters making fine-tunning tricky. For this purpose Implementing learning rate schedule which show cases reducing the learning rate when the validation loss plateaus, to ensure effective training. And other method is to Gradual Unfreezing which means instead of unfreezing all layers at once,starting with the top layers while keeping the lower layers frozen at the beginning, allowing for smoother fine-tunning. Expand the training data and expose the model to more variations [7]. Class Weighting and Hard Example Mining,response to the lesion types imbalance, we used class weights during model training to increase the cost of misclassification of the melanoma class and other rare classes. Hard Example Mining techniques were also used in which a model is trained on harder cases to improve its performance on the more complicated samples. Once the spatial features are extracted with efficientnetb3,the fully connected layer acts as the bridge between these extracted features and and the final classification task. In this before passing the output to the fully connected layer, the multidimensional feature map produced by convolutional layer is flattened into a 1D vector [9]. This transition is essential because the fully connected layer expects 1D input for classification. Generaly, after the fully connected layer, an activation function softmax is used in this process is applied to convert the output into class probabilities, which makes easier to interpret the model's prediction for multiclass classification [4]. In this context of skin lesion detection, the softmax will output probabilities for each lesion detection(e.g.,melanoma,basal cell carcinamo,nevus), and the class with the highest probability is chosen as the final prediction. The fully connected layer also capturing non linear realtionships between the extracted features that the convolutional layers may not have fully captured. By combining features in this dense layer the model can learn different patterns and combinations that distinguish between similar lesion types [2].

To Further refine performance, metrics like precision,recall and F1-score were used, revealing excellent accuracy for certain lesion types Basal cell Carcinamo and Dermatofibroma with 100% accuracy. The main challenges lies in the high morphological similarity between benign and malignant lesions, making accurate classifications difficult. By using most advanced adamax optimizer the model the performance of the model with training phase of 99.95% and validation and testing phase of 94.03% and 93.49% respectively.

II. Literature Survey and Related Works

Mushtaq et al. (2024) integrates deep learning methodologies along with advanced techniques that are used in processing of imagse for improved analysis of skin Lesion images. The study is based on a Convolutional Neural Network(CNN) architecture consists of convolution, pooling, and Layers that are connected fully. Among the major preprocessing techniques involved are hair removal and various filtering methods to improve the dataset of 3,600 images from ISIC database.

The study reports an impressive accuracy of 81.59% for the CNN while a human-machine team showed results as high as 82.95% in multiclass assignments, showing power of human insight combined with AI. There are still some problems with deep network training, among which the disappearing gradient problem stands out and there is the saturation of the activation function, and poor convergence speed of the gradient descent algorithms. This study underlines the possibility of union between image processing and deep learning for heightened accuracy in cancer detection of skin while at the same time addressing some limitations that may exist to affect model performance. [14]

Zhiwei et al. (2023) presented a model that integrates several modules into task of segmentation. The architecture is based on a lightweight transformer block with scale awareness for multiscale characteristics extraction, a deep separable convolutional module for enhancing characters embedding, and an external module for improving effective correlation modeling between samples. It has a multi-scale Feature Aggregation Module that effectively combines the features from multiple scales to enhance performance of model.

It is trained using BceDice loss in a PyTorch environment, with images resized to 256x256 pixels. Data augmentation techniques of various kinds are used for better training diversity. The datasets that have been used here are ISIC2016, ISIC2017, and PH2.

The results prove that Transformer-based Multi-Attention Hybrid Networks(TMAHU-Net) confirms an F1-score of 95.03 on PH2 dataset, where TMAHU-3 fetches 93.30 on ISIC2016 and TMAHU-2 reaches 86.04 on ISIC2017, enhances attention capability of the model with Dilation Gated External Attention(DGEA) module. TMAHU-Net faces some problems such as ambiguous ground truth boundaries, complex lesions might produce a very high false positive rate, and only limited detection of small speckled lesions. These limits indicate further research must be done to develop this model's robustness and accuracy for practical applications. [9]

A new architecture for better accuracy in the skin lesion segmentation task was proposed by Renkai Wu et al. (2023). The authors modified UNet model by adding an HSI module for high-order spatial interaction to effectively capture complex spatial relationships in dermoscopic images to outline detailed structures of lesions.

The ISIC dataset consisting of broad collection of annotated dermoscopic images for purpose of training and validation. The proposed MHorUNet model outperforms conventional segmentation approaches, and results show significant improvement in lesion segmentation with respect to accuracy, considering that model successfully incorporates high-order interactions for processing multi-scale features within images.

The study findings shows model's clinical applicability. MHorUNet is able to provide reliable segmentation outcomes to diagnose skin conditions. In addition,it explores influence of advanced techniques in neural networks for application of medical imaging improves effectiveness of diagnosis tools. [7]

M.A. Rasel et al. (2024)discuss the novel strategy for skin lesion classification while focusing on asymmetry. That is an essential mark, especially in melanomas to be detected during early-stage melanoma detection. Therefore, the approach, called a hybrid model in which geometric pattern analysis can make use of CNN-SVM classification architecture, presents both geometric symmetry-based differences within lesions and how their being asymmetric, or perhaps symmetrical, could mean the difference between having cancer.

The researchers conducted experiments using a geometry-based model wherein asymmetrical patterns achieved a detection rate of 99%. For classification, the architecture would be CNN-SVM-that is, CNN strength used for character extraction by CNN and SVM precision used for better categorization results. High-performance metrics were achieved, including Kappa of 94%, Macro F1-score of 95%, and a weighted F1-score of 97%, indicating that the model is effective at lesion-type classification and assisting in the accuracy of related diagnostic procedures. [1]

Battineni et al. (2022) proposed a hybrid deep learning approach for detecting cancer related to skin and improving accuracy. The approach involves mainly preprocessing with fuzzy logic to identify boundary of lesion, followed by applying morphological operators to remove hair and segment the lesion by using a GrabCut technique. Then, CNNs that are stacked for character extraction are used and an intensified SVM classifier for classifying the lesions by computing the score of features while minimizing overfitting, which improves accuracy.

The model is based on the GC-SCNN algorithm, which was proposed for detecting intricate features in skin cancer . The segmented images go into stacked CNN, ensuring that the extracted features are conveyed through SVM to ensure accurate classification. The outcomes will be of high performance with respect to complex dermoscopic images by correctly differentiating between benign and malignant lesions.

The datasets used involved ISIC2018, ISIC2019, and HAM10000, which offer a broad basis in training and validation. The model achieved sensitivity and specificity of 100%, with an overall reported accuracy in classification of 89.75%. Its performance demonstrated high prediction accuracies in types such as AKIEC, BCC, and BKL. It cuts down processing time by 25-35 milliseconds to achieve a lesion detection time of 2.513 ms. It follows the combination of preprocessing with CNN and SVM enhances diagnosis accuracy and efficiency significantly, therefore offering useful benefits in the area of melanoma detection in medical imaging. [6]

Rout et al. (2023) discussed a hybrid deep learning model for lesion segmentation improvement. This present work proposes a methodology that includes HE, improving the texture in dermoscopic images along with AGCWD. Hair removal techniques are used to clear the artifacts of the images, resizing them to fit the model input size. The accuracy, Jaccard Index, and Dice coefficients are computed as performance metrics by comparing the ground truth images with the generated lesion masks. The work employs the ISIC 2016, 2017, and 2018 datasets, divided into training and testing.

The algorithm enhances through HE and AGCWD techniques, then makes an image datastore out of the processed images and applies segmentation. There were two networks evaluated, ResNet 18 and MobileNetV2, which reached accuracies of 91.98%. Enhanced images showed a remarkable improvement in lesion visibility, although blurring induced by hair removal may sometimes degrade segmentation accuracy. Generally, this work proved that the images enhanced by AGCWD provided equivalent performance with HE-enhanced images on melanoma detection. The hybrid optimization of melanoma detection within automated systems is reached, and the potential of preprocessing, improving the outcomes of segmentation, has been shown. [16]

Ramprasad et al. (2023) proposed a method that involves HAC in combination with RNN-LSTM to segment skin lesions. The approach includes the extraction of spatial features through Conv2D and MaxPooling layers, while it learns about sequential data using Bi-LSTM layers. GIF enhances image quality, and HAC is employed for hierarchical segmentation in order to get an accurate boundary of lesions. Multi-scale feature extraction gives detailed information at each scale and supports the classification model for better accuracy.

It performs preprocessing by using GIF for edge preservation, uses HAC to identify lesion areas, and employs RNN-LSTM for identifying lesions as benign or malignant. With this in place, high performance on the ISBI-2020 dataset, consisting of 128,556 images containing nine types of skin diseases, can be expected, having recorded an accuracy of 90.71% in segmentation and 89.46% in classification. In the comparative tests, better results were obtained compared to KNN, SVM, and CNN, with significant accuracy reduction without GIF and HAC. This multilevel and structured approach of the method promises robust lesion detection in a clinical setting. [15]

Anisi et al. (2023) proposed a hybrid approach to skin cancer lesion detection. It combines fuzzy logic-based image segmentation with the modified version of deep learning YOLO models. Fuzzy Segmentation algorithm enhances dermoscopic images, while convolutional layers and residual connections across YOLO improve accuracy. Adoption of multi-scale feature concatenation techniques for better feature extraction can be done.

ISIC 2017, ISIC 2018, PH2, and HAM10000 datasets are comprised of 10,015, 2,000, and numerous RGB images, respectively. On PH2, the sensitivity was 95%, the specificity was 96.25%, and the IoU was 94%. In comparison, this modified version of the YOLO model was much faster and highly accurate in its classification compared to standard classifiers. Moreover, the modification further showed its robustness for both detection and segmentation tasks on different datasets. [18]

Naveed Ahmad et al. (2023) targetted skin lesion classification using the benchmark datasets HAM10000 and ISIC2018. The procedure relies on augmented data in sequence to enhance quality and improve the size of the training dataset. It relies on transfer learning by taking models that are trained previously such as Xception and ShuffleNet. Feature extraction will be carried out with the help of a global average pooling layer; afterward, feature fusion is to be done by adopting a Serial-Threshold approach. Thereafter, BOA will go for feature selection in such a manner that it will reduce dimensionality by emulating the butterfly foraging behavior.

The obtained results from the framework are accuracy-91.3% for the ISIC2018 dataset and 91.5% for the HAM10000 dataset. Performance evaluation was done using ten classifiers, including Neural Networks and SVMs, which ensured better accuracy and computational time reffered to existing methods. Herein, this work underlined the effectiveness of explainable AI integrated with deep learning techniques for robust skin lesions that are related to skin recognition and enhancement of the accuracy with interpretability. [2]

Faruq Aziz et al. (2024) presented advanced skin lesion detection based on the YOLOv9 architecture with enhanced features by the use of Programmable Gradient Information (PGI) and a Generalized Efficient Layer Aggregation Network (GELAN). This work has used 2721 training images, 288 validation images, and 145 test images divided into six distinct classes of skin diseases. These results include a precision of 60.5%, a recall of 86.0%, and MAP of 81.4%. In the case of acne lesions, high precision was achieved at 85.6% and a recall of 96.7%, though there were some difficulties with detecting atopic dermatitis and psoriasis. Limitations of this study are that the dataset size is generally small and may result in biases within performance. The training was done for 30 epochs with a batch size of 16, meaning focused effort on optimization concerning the detection capabilities, yet acknowledging the constraint of the data employed for this work. [3]

Rachana et al. (2024) introduced a framework: the integration of JAEO with the LeNet architecture for skin cancer detection. The inspiration for this novelty was taken from a preprocessing step of bilateral filtering that removed all noise to bring out high-quality images. Later, skin lesion segmentation is done via TransUNet, which is very effective in segmenting the regions of interest within dermoscopic images. The authors further applied techniques that augment the data to improve variability of dataset training and enabled better generalization in feature extraction. Then, the LeNet model takes advantage of these features to make accurate predictions about skin cancer presence. An outstanding accuracy of 91.99%, with sensitivity and specificity rates at 90.95% and 92.13%, respectively. This level of performance demonstrates that JAEO-LeNet model is both reliable and superior to traditional methods of detection.

The large SIIM-ISIC Melanoma classification database used for this study was composed of 33,126 dermoscopic images. This large dataset allows comprehensive testing of model's performance for various skin lesion types and strengthens the argument of using big data in the development of strong diagnostic tools. Results show that optimized deep learning techniques can change dermatological practices and may contribute significantly to early detection of skin cancer by improving patient outcomes in a clinical setting. [4]

The author Halit Çetiner (2024) proposed a deep learning model using transfer learning techniques for classifying skin cancers according to their types. The model focuses on seven skin cancer types, using the pre-trained DenseNet201 architecture on the ImageNet dataset, which massively improves the model's predictiveness.

The model that was proposed attained an accuracy of 92.51% and outperformed basic DenseNet201 by 15%. It provides excellent metrics of performance: F1 score, recall, and precision have shown high classification capability. ISIC 2018 dataset used for training. It allows the model to learn from an exhaustive set of dermoscopic images effectively. The average accuracy of the model is reported to be 91.73%, with its potential reliable skin cancer diagnostics. The focus is mainly on how deep learning methodologies, and specifically transfer learning, improve accuracy in identification and analysis of skin lesions. The current work supports the growing trend in integrating AI and machine learning technologies into medical imaging for improved diagnostic performance. [21]

Bhuvaneswari et al. (2022) worked on HAM10000 dataset of 10,015 images of dermoscopic lesions utilizing machine learning and CNNs. It performs data augmentation and K-fold cross-validation. Other techniques used in the approach involve resizing images for reduced memory as well as latency alongside global feature descriptors to enhance feature extraction efficiency. CNNs were applied based on set hyperparameters, with final layer designed as a Softmax layer for determination of classification for lesions. As comparisons, the Random Forest algorithm of machine learning is applied and, in some situations, better than most other approaches. CNN reached an accuracy of 91.18%, whereas Random Forest enhanced performance metrics such as recall, precision, and F1-score. Extensive experiments verified low model loss with minimal overfitting post-training. However, study found that there are strong class imbalances in dataset, which affect the performance of model even after performing data augmentation to reduce such imbalances. Moreover, the study found that computational efficiency of models is poor and requires proper validation for reliable diagnostic results in clinical applications. [17]

Muhammad Attique Khan et al. (2024) proposed an approach consists of an application of augmentation of data methods, such as rotating images and flipping images. It utilizes pre-trained models, namely DarkNet-19 and MobileNet-V2, which are further optimized by employing Firefly Algorithm and hybridized to refine feature selection in heuristic methods for maximum precision in classification of skin lesions. Feature vectors are combined through serial and concatenation approaches, and classifiers like SVM and CNN are used to classify seven different classes of skin in the ISIC 2018 dataset. The study achieved an accuracy of 89.0% with Cubic SVM achieving a high accuracy of 87.5%. Precision and sensitivity rates were also strong, contributing to an F1 score of 87.45%. These improvements suffer from the use of an unbalanced dataset for model accuracy, along with very high computational requirements. These models may tend towards overfitting, meaning generalization may decrease from varying types of lesions the models are not trained for, so these results point toward good generalization when this particular method of deep learning techniques would yield high improvements in domain of medical image classification and refinement would better allow to enhance generalizability of these models with reduction of their resource intensity. [12]

Hosny et al. (2024) presented this model is found to counter the problems in the skin lesion classification with degradation inside the image and limited dataset; hence, the accuracy of the model is seen to be 92.1%, sensitivity of 98.9%, specificity of 94.17%. The proposed system outperforms all the other models that have been so far applied for the purpose of skin cancer diagnosis, but this model is too computationally heavy and huge sets are required for the training process. Actually, its depth will bring more problems with deterioration, as it enhances classification capability; however, improved efficiency as well as accuracy of classification for skin lesion is identified in this work, which still hints toward the presence of potential optimisations inside future applications of this network. [10]

III. Methodology

A. Description of Dataset

This dataset for the research study is sourced from Kaggle. The dataset consists of all the relevant diagnostic categories for 10,015 images of dermoscopic pigmented lesions. There are 10,015 dermatoscopic images obtained from a diverse group of people, acquired and stored in various imaging modalities suitable for academic purposes of machine learning, mainly multi-label classification of skin lesions. The set consists of a broad spectrum of diagnostic categories pertinent to pigmented skin lesions, ensuring an appropriate representation across all major types: akiec, bcc, bkl, df, mel, nv, and vasc. Over 50% of the cases are confirmed by histo. In summary, the above dataset provides a strong ground truth while the rest of them are confirmed by further studies, expert agreement or in-vivo confocal microscopy (confocal). Such a diverse dataset of such quality thus could turn out to be a treasure in developing transfer learning methods applicable in dermatological image classification. Different diagnostic classes should be covered and thus catered for better understanding pigmented lesions.

B. Data Preprocessing

Data Preprocessing is the most crucial part of deep learning models. In this research, many preprocessing techniques are used to improve the accuracy of the model such as Removing the noise in the images and handling the class imbalance. The images in the dataset have hair artifacts. The first step is to

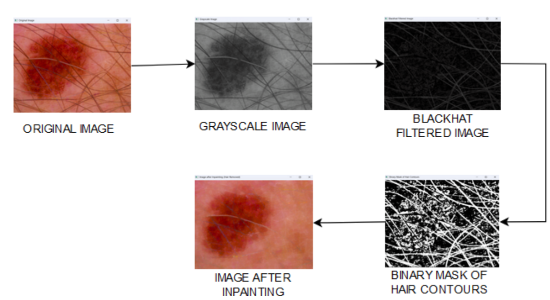

Fig. 3. Removing the Hair artifacts from the images

remove the hair artifacts from the images. for this, a technique in computer vision called Blackhat Morphological Filtering and Inpainting is used. There are a few steps in this technique. This technique starts with converting the original image to grayscale, which simplifies the focus on intensity variations, and then a blackhat filter is applied to the grayscale image to highlight dark features like hair contours against the lighter skin background. After that, the image undergoes thresholding to create a binary mask that accurately isolates hair regions. This binary mask is then used in an inpainting operation, where surrounding pixels fill in the masked areas, effectively removing the hair artifacts. Figure 3 shows the steps present in removing the hair artifacts(noise) from the images.

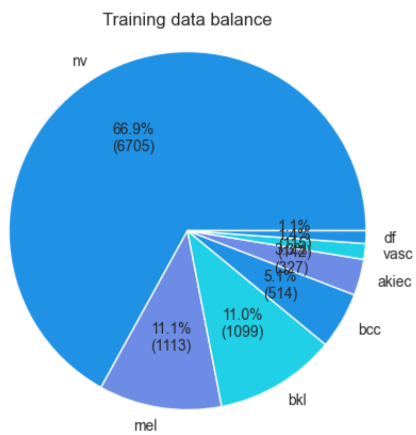

Fig. 4. distribution of images into different classes

After removing the noise in the dataset the next step is to handle a class imbalance. Figure 1 represents a pie chart that illustrates the distribution of different classes within a training dataset for skin lesion image classification. The melanocytic nevi class dominated the other classes by 66.9% and the dermatofibroma class is less dominant with a percentage of 1.1%. To handle this class imbalance, there is a technique called image data augmentation. This technique creates new images in the minority class, and another technique called downsampling, which is a specific type of data sampling method where, it reduces the number of samples from the majority class to match the number of samples in the minority class, balancing the dataset for improved model performance.

1) Data augmentation: Another technique of data enhancement quite common with deep learning is data augmentation. That has more data in training. Deep learning methods will work better if the volume of training data is increased. The dataset consists of seven classes of skin called akiec, bcc, bkl, df, mel, nv and vasc have 327, 514, 1099, 115, 1113, 6705, and 142 images respectively in the dataset training. This difference in images causes the decreased accuracy of the model. So, we will try to increase the minority classes by augmentation of image data. New images are created by performing some operations like rotating the images, shifting the images to enable the model to learn positional flexibility, zooming the image, flipping the images, either horizontally or vertically, and changing the brightness level of the images. Figure 5 shows some of the augmented images like horizontal and vertical flips of the images as well as rotation of images. This makes the model robust.

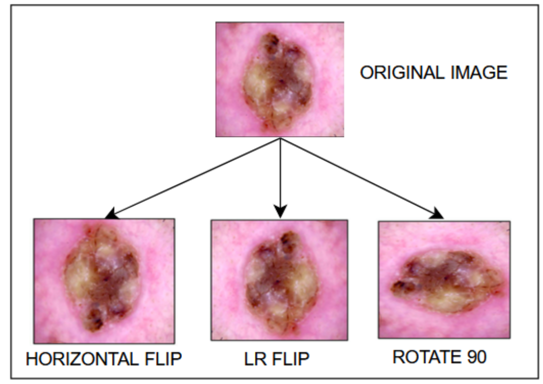

Fig. 5. Augmentated images of minority classes

C. *EfficientNetB3 Algorithm*

EfficientNetB3: This is one of the most famous convolutions in a neural network, developed by Google researchers in 2019. It uniformly scales the depth, width, and resolutions of model dimensions to offer improved accuracy and efficiency systematically. In this model compound scaling is used that employed compound coefficient ϕ for uniform scaling of model dimensions, where the help of compound coefficients calculates Network Scaling Dimensions- Width Scaling Coefficient, Depth Scaling Coefficient, Resolution Scaling Coefficient, which are

$$\begin{aligned} \text{depth: } d &= \alpha^{\phi} \\ \text{width: } w &= \beta^{\phi} \\ \text{resolution: } r &= \gamma^{\phi} \end{aligned} \tag{1}$$

where

- α, β, γ are constants
- ϕ is a compound coefficient

The above constants should follow the constraint as given below

$$\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2, \qquad \alpha \geq 1, \beta \geq 1, \gamma \geq 1 \tag{2}$$

Different versions of EfficientNet (B0-B7) are defined using these compound scaling coefficients, and EfficientNetB3 is defined by applying a depth multiplier of 1.4, a width multiplier of 1.2, and a resolution of 300x300 pixels to the base model. Figure 6 represents the layers present in the Baseline Network of EfficientNet and uses the Mobile Inverted Bottleneck layer and convolutional layer as their building blocks.

1) Convolution layer: A convolutional layer is the most important layer in convolutional neural networks. It contains multiple filters and applies a filter to an input image, performing element-wise multiplication and summation to produce a feature map that highlights specific features like edges or textures.

Fig. 6. Baseline Network of EfficientNet

2) Mobile Inverted Bottleneck Convolution: A Mobile Inverted Bottleneck Convolution is a powerful convolutional layer that balances between the computational efficiency and model performance. This is one of the building blocks for many state-of-the-art neural networks, including the EfficientNet. Mobile Inverted Bottleneck Convolution: uses the depth-wise separable convolutions as the basic principle. This breaks the convolutional operation into two separate operations: one is a depth-wise convolution and another is a point-wise convolution. In depth-wise convolution, a single filter will be applied to each input channel individually, which reduces the number of computations and parameters in general. After then, the point-wise convolution is used to expand the feature space, enabling the network to learn complex patterns. By using this inverted residual structure, Mobile Inverted Bottleneck Convolution layers can achieve significant computational savings compared to traditional convolutional layers. Additionally, Mobile Inverted Bottleneck Convolution layers can be stacked together to form deeper networks, leading to improved performance in model accuracy.

D. Fully Connected Neural Network

Neural networks are the most popular machine learning algorithms, which are capable of learning complex patterns and relationships between the data. For classifying images into lesions, implemented a fully connected feed-forward neural network and this network is used as a Classifier for the model. Each neuron in layers performs simple computations, recognizes the complex associations between the data, and activation functions are used to determine the output of neurons. Some of the most commonly used functions are Linear, Softmax, Signmoid, and ReLU. Here the model used the Softmax activation function for classification.

1) Rectified linear unit:
A rectified linear unit is a non-linear activation function that transforms the negative values to zero and positive values to itself.

$$y = (W * x + b) \tag{3}$$

$$ReLU(y) = max(0, y) \tag{4}$$

where:

- x represents the input vector of the neuron
- W is the weight vector of the particular neuron
- b is a bias of the particular neuron
- ReLU is the rectified linear unit activation function

This Activation function is used in layers other than the output layer in fully connected neural network.

2) Softmax function:
A softmax function is a non-linear function used particularly in the output layer for multi-class classification problems. It converts calculated weights into a probability distribution over multiple classes.

$$S(z_i) = \frac{e^{z_i}}{\sum_{i=1}^{n} e^{z_i}} \tag{5}$$

where

- e is Euler's number
- z_i are the calculated weights of neurons
- $S(z_i)$ are probability distribution over multiple classes.

The softmax function is used in the output layer of the neural network model and it helps to find the class label for a given input.

E. Model Hyperparameters

Model hyperparameters are crucial for model accuracy, so tuning the hyperparameters is very much important. After several experiments, parameters are selected. Table I shows the parameters used in training the EfficientNet Model. Adaptive Moment Estimation (Adam) optimizer is the best case for the present dataset. This optimizer combines the strengths of momentum and the RMSprop algorithm. In this case, categorical cross-entropy has been used as a Loss Function. A loss function measures an error value between calculated values from the machine learning model and ground truths. Different types of loss functions include Mean Squared Error and cross-entropy Loss. selection of loss function depends upon what kind of data and the present dataset is categorical so, used Categorical cross-entropy.

TABLE I
EFFICIENTNET MODEL HYPERPARAMETERES

Hyperparameter	Value
optimizer	adam
Loss function	Categorical cross-entropy
Epochs	38
Batch Size	32
Learning Rate	0.00001

F. Model Architecture

The model Architecture is presented in figure 7. In this research, the most popular technique called Transfer Learning is used. It is a technique based on ML, that allows the reprocess the models that are trained previously as a base model for developing models for new related tasks. In this research, EfficientNetb3 is pre-owned, which is trained on the ImageNet dataset and the model's learned features from the ImageNet dataset are used to drqwout important characters from the present skin lesion dataset. This is how transfer learning is important for achieving state-of-art accuracy. The steps present in the proposed model are, the first step is to remove the hair artifacts from the images and then convert

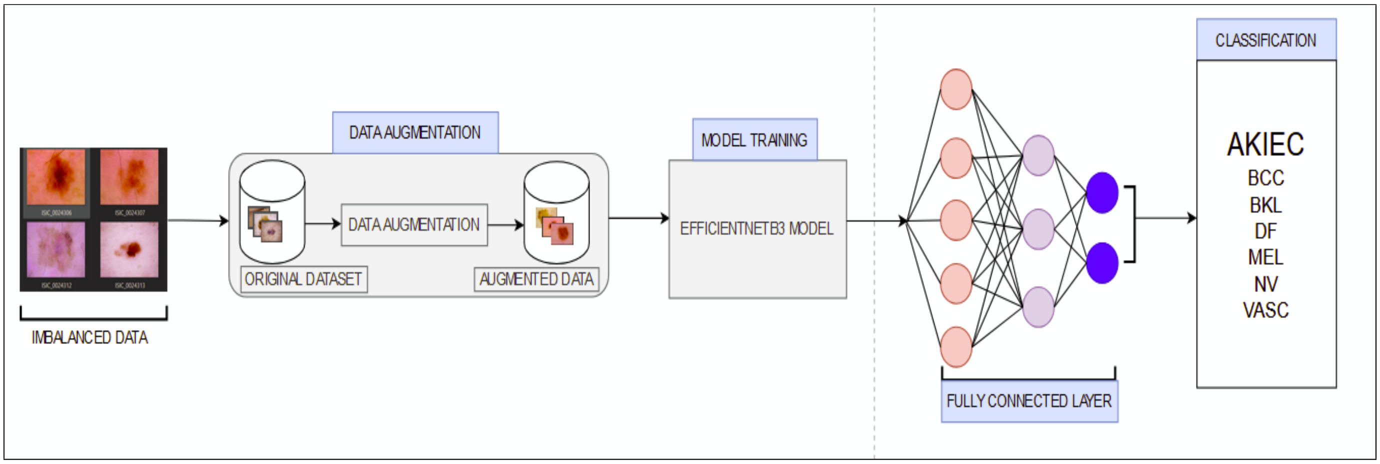

Fig. 7. Architecture Diagram

the class-imbalanced dataset into a balanced dataset using the image data augmentation technique. Then the updated dataset is given to the model for training, and then after learned parameters are given to the fully connected Neural Network, which classifies the skin lesion images into respective classes.

IV. Results

A. Confusion Matrix

A confusion matrix is a basic tool utilize to assess the performance of machine learning and statistics classifiers by comparing the predicted labels and actual (true) labels. In this matrix, four such important outcomes are measured; True Positives, correct predictions of positive cases; TN, correct prediction of negative cases; FP, or false predictions of positive cases (which are Type I errors); and FN, the incorrect predictions of negative cases or Type II errors. It helps make such matrices valuable because matrices serve to help in assessing how practical the implications of model errors really are. Model errors can bring different kinds of implications in different fields, even as sensitive as medical diagnoses.

Figure 8 represents a confusion matrix of seven different categories like akiec, bcc, bkl, df, mel, nv, and vasc diseases. The diagonal elements represent the correct classification, which is known as true positives. Off-diagonal elements depict misclassifications. In this model, some categories performed exceptionally well with 100% accuracy in classifying bcc, df, and vasc diseases. Akiec performed with good strength at 99.07% with minimum misclassification. However, it has some notable confusion in other categories. bkl category shows 83.33% accuracy and has some misclassifications, especially with mel (9.17%) and nv (4.17%). the mel category achieved 87.39% accuracy, but it shows some confusion with nv (8.11%), and it reached 85.71% accuracy with some misclassifications as mel (8.33%) and bkl (4.76%). These values shows the performance of the proposed model.

Fig. 8. confusion matrix

B. Evaluation metrics

precision, recall, and F1-score are three critical measurements that provide comprehensive insights into model performance. These metric are defined as

Precision: Precision defines the number of true positive observations in relation to the predicted positive observations.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (6)$$

Recall(Sensitivity): Recall or sensitivity or true positive rate is the proportion of positive prediction that was accurately realized to total actual positives. Recall is most important when a false negative is expensive.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{7}$$

TABLE II
PRECISION, RECALL, AND F1-SCORE FOR EACH CLASS

Class	Accuracy	Precision	Recall	F1-score
akiec	99.97%	0.98	0.99	0.99
bcc	100%	0.97	1.00	0.98
bkl	83.33%	0.89	0.83	0.86
df	100%	0.98	1.00	0.99
mel	87.39%	0.80	0.87	0.83
nv	85.71%	0.91	0.86	0.88
vasc	100%	1.00	1.00	1.00

F1-Score: A balanced measure between precision and recall is the F1-score, which can be understood as the unique and relative mid of via-precision and jets. It is especially helpful in scenarios where both false positives and false negatives are underestimated and it is essential to get a single measurement integrating both precision and recall in case of class imbalance. Formula:

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{8}$$

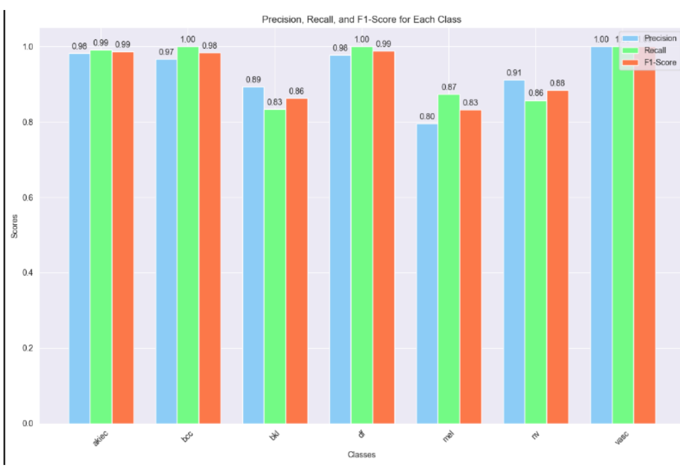

Fig. 9. Evaluation metrics of Seven diseases

The bar graph presented in figure 9 illustrates Evaluation metrics across seven different classes. The visualization shows that classes vary in performance levels with some classes performing excellently and others showing good performance but a little lower. For instance, the bcc, df, and vasc classes show excellent performance with perfect or nearly perfect scores in all the metrics, which implies that they have highly reliable classification capabilities. The Akiec class shows excellent performance since its metrics are constantly ranging from 0.98-0.99, which is a very dependable classification. Other classes are bkl, mel, and nv, with good performance and some potential for improvement; the metric varies from 0.80 to 0.91. Pattern analysis over all three metrics for most classes reflects an almost balanced performance. The variation of precision over recall for any class would suggest some strategic trade-off in this model's prediction approach, and the F1-score typically aligns well with both precision and recall and confirms balanced overall performance. The model is well-tuned as it presents reliable performance for all classes. Medical diagnosis applications would highly care about this since the failure to either recognize true positivity or to exclude a pathology can cause critical effects on patient treatment.

C. Discussion:

TABLE III
PROPOSED MODEL COMPARISON WITH RECENT EXISTING MODELS ON HAM10000 DATASET

Existing work versus proposed work	Accuracy (%)
Proposed EfficientNet Model	**93.46**
InceptionResNetV2	88.20
ResNetXt101	87.20
Xception	86.47
ResNet50, ResNet101+kPCA+SVM RBF	89.80
24-layered CNN	86.50
MobileNetV2-LSTM	85.34
EW-FCM+wide-shufflenet	84.80
Modified MobileNet	83.93
Shifted MobileNetV2	81.90
Vgg16+googLeNet ensemble	81.50
9-layered CNN	80.00

Our proposed EfficientNet-based model reached an accuracy of 93.46% on the HAM10000 dataset, surpassing many advanced architectures, such as InceptionResNetV2 and ResNetXt101. High performance is an indication that the balance between computational efficiency and classification accuracy given by EfficientNet gives it an edge in processing and analyzing dermoscopic images where fine-grained detail is crucial for accurate lesion classification.

ResNetXt101 and InceptionResNetV2 are the progressive models in the comparison with ResNet and Inception. Though your model has not been trained with better parameter tuning, improving architecture design of EfficientNet has helped your model surpass the above two models by achieving accuracies at 88.20%. It may be due to compound scaling strategy for depth, width, and resolution by EfficientNet while optimizing for performance on the data of images.

Xception Competitive Performance: Another very powerful model is Xception that achieved an accuracy of 86.47%, however, it tends to be computationally heavy. Therefore, your comparable performance from the EfficientNet model likely at a reduced computational cost, indicates your model efficiency in handling such high resolution medical images ideal for real-time applications in healthcare.

CNN ensembles and old models give low performance Lower performance than that given by 9-layered CNN, VGG16+googLeNet ensemble scored less than 85% to show how a basic CNN architecture, or ensembles without even highly-tuned techniques, could fail at describing the details within the complex features arising in images of skin lesion. That's why, here, EfficientNet shines ahead.

Role of Transfer Learning and LSTM Combinations: MobileNetV2-LSTM (85.34%) uses LSTM layers since it learns sequences. It captures the spatial dependencies well but does not top the EfficientNet on this dataset. The streamlined design of EfficientNet will bring a better feature extraction without the use of sequential layers through its balanced accuracy and simplicity.

The proposed EfficientNet model achieves a state-of-the-art design optimized by both feature extraction and classification in medical images, notably skin lesions. With significant accuracy on the HAM10000 dataset, it sits comfortably within the top-tier models for possible clinical application in dermatology or earlier diagnosis of melanoma.

D. Conclusion:

In conclusion, the proposed approach is a holistic solution to the prevalent issue of data imbalance in medical image classification, especially in skin session diagnosis. Data imbalance is a big challenge in this domain because certain skin conditions are much rarer than others, resulting in fewer images available for training in those classes. This can lead to deep learning models biasing toward more frequently repeated classes. Then, the predictive accuracy decreases when making predictions on conditions that are less commonly represented. This has led to the need to use a key technique for addressing data augmentation techniques, which means adding extra samples that are developed from the dataset through techniques of rotation, flipping, zooming, and shifts. These transformations enhance the richness of the dataset reveal to the model all perspectives on each class and facilitate better learning across classes with fewer samples.

As a powerful as well as efficient option, this classification model takes up the base model of EfficientNet-B3. This kind of architecture of the model is very high resolutions. In most cases, for the proper classification or diagnosis of a skin lesion, its features must be extracted with high resolutions. One variant in that family is a mid-sized; it can balance, relatively speaking, the two above-mentioned properties, given their computational costs, this seems more suitable for the deployment scenarios that might happen in very scarce resource contexts, including possibly medical use cases and potential deployment domains.

Once that augmented set has been input into an EfficientNet-B3, it should follow a few convolution layers developed to pick up harder or more complicated patterns in the images. This is an important ability to differentiate between different types of skin conditions that can appear the same but have various clinical implications. Conditions like AKIEC, BCC , BKL , DF , MEL , NV , and VASC require high feature sensitivity for proper classification. The fully connected layers on the model architecture also provide discriminative features with which the model can output very accurate and subtle predictions for these classes.

In summary, this pipeline demonstrates a very robust methodology that achieves a very balanced and accurate classification for medical imaging, effectively correcting class imbalance while optimizing performance. This helps improve diagnostic accuracy and reduce bias; more importantly, it will also help clinicians have an efficient automatic analysis tool for the determination of skin lesions. Adding to this are the approaches combining data augmentation with a more complex architecture. Therefore, dermatology can help streamline the entire process of diagnosis while improving detection along with patient outcomes. This solution illustrates how deep learning and data augmentation can be integrated into developing a balanced, high-performance system for medical image classification to solve critical challenges in the field and support more equitable and accurate clinical decision-making.

References

[1] Asymmetric lesion detection with geometric patterns and cnn-svm classification. *Computers in Biology and Medicine*, 179:108851, 2024.

[2] Hafiz Ahmad, Jamal Shah, Muhammad Khan, Jamel Baili, Ghulam Ansari, Usman Tariq, Ye Kim, and Jaehyuk Cha. A novel framework of multiclass skin lesion recognition from dermoscopic images using deep learning and explainable ai. *Frontiers in Oncology*, 13:1151257, 06 2023.

[3] Faruq Aziz and Daniati Saputri. Efficient skin lesion detection using yolov9 network. *Journal Medical Informatics Technology*, pages 11–15, 03 2024.

[4] Rachana Babu and Felix Philip. Optimized deep learning for skin lesion segmentation and skin cancer detection. *Biomedical Signal Processing and Control*, 95:106292, 09 2024.

[5] Naliniprava Behera, Akhilendra Pratap Singh, Jitendra Kumar Rout, and Bunil Kumar Balabantaray. Melanoma skin cancer detection using deep learning-based lesion segmentation. *International Journal of Information Technology*, pages 1–16, 2024.

[6] Usharani Bhimavarapu and Gopi Battineni. Skin lesion analysis for melanoma detection using the novel deep learning model fuzzy gc-scnn. *Healthcare*, 10:962, 05 2022.

[7] Qing Chang. Mhorunet: High-order spatial interaction unet for skin lesion segmentation. 11 2023.

[8] Priya Choudhary, Jyoti Singhai, and JS Yadav. Skin lesion detection based on deep neural networks. *Chemometrics and Intelligent Laboratory Systems*, 230:104659, 2022.

[9] Zhiwei Dong, Jinjiang Li, and Zhen Hua. Transformer-based multi-attention hybrid networks for skin lesion segmentation. *Expert Systems with Applications*, 244:123016, 12 2023.

[10] Khalid M. Hosny, Wael Said, Mahmoud Elmezain, and Mohamed A. Kassem. Explainable deep inherent learning for multi-classes skin lesion classification. *Applied Soft Computing*, 159:111624, 2024.

[11] Ayushi Jain, Neha Mittal, Madasu Hanmandlu, and Arvind Pandey. Skin lesion detection: An analysis of advanced computational approaches. In *Developments Towards Next Generation Intelligent Systems for Sustainable Development*, pages 204–224. IGI Global, 2024.

[12] Muhammad Khan, Ameer Hamza, Dr. Mohammad Shabaz, Seifeine Kadry, Saddaf Rubab, Muhammad Bilal, Muhammad Akbar, and Kesavan Suresh. Multiclass skin lesion classification using deep learning networks optimal information fusion. *Discover Applied Sciences*, 6, 05 2024.

[13] Yuexiang Li and Linlin Shen. Skin lesion analysis towards melanoma detection using deep learning network. *Sensors*, 18(2):556, 2018.

[14] Snowber Mushtaq and Omkar Singh. Implementing image processing and deep learning techniques to analyze skin cancer images. pages 2210–142, 03 2024.

[15] M.V.S. Ramprasad, S.S.V. Nagesh, V. Sahith, and Rohith Lankalapalli. Hierarchical agglomerative clustering-based skin lesion detection with region based neural networks classification. *Measurement: Sensors*, 29:100865, 07 2023.

[16] Ranjita Rout, Priyadarsan Parida, and Sonali Dash. Automatic skin lesion segmentation using a hybrid deep learning network. 15:238–249, 06 2023.

[17] Bhuvaneshwari Shetty, Roshan Fernandes, Anisha Rodrigues, Rajeswari Chengoden, Sweta Bhattacharya, and Kuruva Lakshman. Skin lesion classification of dermoscopic images using machine learning and convolutional neural network. *Scientific Reports*, 12, 10 2022.

[18] Sumit Singh, Vahid Abolghasemi, and Hossein Anisi. Fuzzy logic with deep learning for detection of skin cancer. *Applied Sciences*, 13:8927, 08 2023.

[19] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019.

[20] Sulaiman Vesal, Shreyas Malakarjun Patil, Nishant Ravikumar, and Andreas K Maier. A multi-task framework for skin lesion detection and segmentation. In *International Workshop on Computer-Assisted and Robotic Endoscopy*, pages 285–293. Springer, 2018.

[21] Halit Çetiner. Skin lesions identification and analysis with deep learning model using transfer learning. *Osmaniye Korkut Ata Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 7:1030–1045, 06 2024.